

[DASHBOARD](#) / [I MIEI CORSI](#) / [APPELLI DI CLAUDIO SARTORI](#) / [SEZIONI](#) / [MACHINE LEARNING](#) / [MACHINE LEARNING THEORY](#)

Iniziato	Thursday, 13 January 2022, 15:13
Stato	Completato
Terminato	Thursday, 13 January 2022, 15:36
Tempo impiegato	23 min. 5 secondi
Punteggio	15,00/15,00
Valutazione	30,00 su un massimo di 30,00 (100%)

Domanda **1**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main reason for the *MinMax scaling* (also known as "*rescaling*") of attributes?

Scegli un'alternativa:

- ☐ a. Map all the numeric attributes to the same range, in order to prevent the attributes with higher range from having prevalent influence
- ☐ b. Map all the nominal attributes to the same range, in order to prevent the values with higher frequency from having prevailing influence
- ☐ c. Change the distribution of the numeric attributes, in order to obtain gaussian distributions
- ☐ d. Remove abnormal values

Domanda **2**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following *is not* an objective of feature selection

Scegli un'alternativa:

- a. Select the features with higher range, which have more influence on the computations
- b. Reduce time and memory complexity of the mining algorithms
- c. Reduce the effect of noise
- d. Avoid the *curse of dimensionality*

Domanda **3**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What is the *single linkage*?

Scegli un'alternativa:

- › a. A method to compute the separation of the objects inside a cluster
- › b. A method to compute the distance between two classes, it can be used in decision trees
- › c. A method to compute the distance between two sets of items, it can be used in hierarchical clustering
- › d. A method to compute the distance between two objects, it can be used in hierarchical clustering

Domanda **4**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the definitions below:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Which of the formulas below computes the *precision* of a binary classifier?

Scegli un'alternativa:

- a. $(TP + TN) / (TP + FP + TN + FN)$
- b. $TN / (TN + FP)$
- c. $TP / (TP + FN)$
- d. $TP / (TP + FP)$

Domanda **5**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What is the *cross validation*

Scegli un'alternativa:

- a. A technique to improve the quality of a classifier
- b. A technique to improve the speed of a classifier
- c. A technique to obtain a good estimation of the performance of a classifier when it will be used with data different from the training set
- d. A technique to obtain a good estimation of the performance of a classifier with the training set

Domanda **6**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

A Decision Tree is...

Scegli un'alternativa:

- a. A tree-structured plan of tests on single attributes to forecast the cluster
- b. A tree-structured plan of tests on single attributes to forecast the target
- c. A tree-structured plan of tests on multiple attributes to forecast the target
- d. A tree-structured plan of tests on single attributes to obtain the maximum purity of a node

Domanda **7**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main purpose of *smoothing* in Bayesian classification?

Scegli un'alternativa:

- a. Dealing with missing values
- b. Classifying an object containing attribute values which are missing from some classes in the test set
- c. Classifying an object containing attribute values which are missing from some classes in the training set
- d. Reduce the variability of the data

Domanda **8**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following clustering methods is **not** based on distances between objects?

Scegli un'alternativa:

- ☐ a. DBSCAN
- ☐ b. Hierarchical Agglomerative
- ☐ c. Expectation Maximization
- ☐ d. K-Means

Domanda **9**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following characteristic of data can reduce the effectiveness of K-Means?

Scegli un'alternativa:

- ☐ a. Presence of values with high frequency
- ☐ b. All the variables have the same distribution of values
- ☐ c. All the variables are the same range of values
- ☐ d. Presence of outliers

Domanda **10**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the statements below is true? (One or more)

Scegli una o più alternative:

- a. DBSCAN can give good performance when clusters have concavities
- b. Increasing the radius of the neighbourhood can decrease the number of noise points
- c. Sometimes DBSCAN stops to a configuration which does not include any cluster
- d. DBSCAN always stops to a configuration which gives the optimal number of clusters

Domanda **11**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What is the meaning of the statement: "*the support is anti-monotone*"?

Scegli un'alternativa:

- a. The support of an itemset is always smaller than the support of its supersets
- b. The support of an itemset never exceeds the support of its subsets
- c. The support of an itemset never exceeds the support of its supersets
- d. The support of an itemset is always smaller than the support of its subsets

Domanda **12**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

How does *pruning* work when generating frequent itemsets?

Scegli un'alternativa:

- a. If an itemset is frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated
- b. If an itemset is not frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated
- c. If an itemset is frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated
- d. If an itemset is not frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated

Domanda **13**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the effect of the *curse of dimensionality*

Scegli un'alternativa:

- a. When the number of dimensions increases the classifiers cannot be correctly tuned
- b. When the number of dimensions increases the euclidean distance becomes less effective to discriminate between points in the space
- c. When the number of dimensions increases the results tend to be prone to overfitting
- d. When the number of dimensions increases the computing power necessary to compute the distances becomes too high

Domanda **14**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

When is polynomial regression appropriate?

- a. When there is more than one predicting attribute
- b. When the relationship between the predicting variable and the target cannot be approximated as linear
- c. When it is necessary to project the data into a higher dimensional space
- d. When the target values are not linearly separable

Domanda **15**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What does K-means try to minimise?

Scegli un'alternativa:

- a. The *distortion*, that is the sum of the squared distances of each point with respect to the points of the other clusters
- b. The *separation*, that is the sum of the squared distances of each point with respect to its centroid
- c. The *separation*, that is the sum of the squared distances of each cluster centroid with respect to the global centroid of the dataset
- d. The *distortion*, that is the sum of the squared distances of each point with respect to its centroid

Vai a...

[Data Mining Theory](#) ►